P517317.PDF [Page 3 of 35]

DEFENCE R&D DÉFENSE

Assessing intelligent software agents for training maritime patrol aircraft crews

Stuart C. Grant

Defence R&D Canada

Technical Report

DCIEM TR 2001-036

December 2001

National Defence / Défense nationale

Canada

Assessing intelligent software agents for training maritime patrol aircraft crews

Stuart C. Grant

Defence and Civil Institute of Environmental Medicine

Technical Report

DCIEM TR 2001-036

December 2001

517317.PDF [Page: 5 of 35]

Author

Stuart C. Grant, Ph.D.

Approved by

Lochlan E. Magee, Ph.D.

Head, Simulation and Modelling for Acquisition, Rehearsal, and Training Section

Approved for release by

K.M. Sutton

Document Review and Library Committee

This study, approved by the Defence and Civil Institute of Environmental Medicine Human Research Ethics Committee, was conducted in conformity with the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans.

The information contained herein has been derived and determined through best practice and adherence to the highest levels of ethical, scientific, and engineering investigative principles. The reported results, their interpretation, and any opinions expressed therein, remain those of the authors and do not represent, or otherwise reflect, any official opinion or position of DND or the Government of Canada.

© Her Majesty the Queen as represented by the Minister of National Defence, 2002

© Sa majesté la reine, représentée par le ministre de la Défense nationale, 2002

104-05480/

Abstract

Training simulators often require the participation of several people to play the role of supporting players in the simulated operation. Use of intelligent software agents to play the role of these personnel has the potential to reduce support staff and increase an instructor's control of training. This report evaluates a simulator prototype developed for the CP140 Aurora maritime patrol aircraft that incorporated intelligent software agents to play the roles of the Tactical Navigator and an Acoustic Sensor Operator. Human crews, intelligent agent crews, and mixed human-agent crews performed a simulated antisubmarine mission. Mission performance and crew communications were recorded and rated to determine whether the intelligent software agents could perform individual crewmember functions and whether they could provide the interaction necessary for crew coordination training. The results indicate that (1) agents can perform individual crewmembers' functions; (2) agent interaction with humans is sufficient to allow humans to perform their own tasks; and (3) the agents did not interact in a way suitable for crew coordination training. It is concluded that the prototype is suitable for supporting individual training, but the agents' knowledge base must explicitly address team dynamics if crew coordination training is to be supported.

Résumé

Les simulateurs d'entraînement font souvent appel à plusieurs personnes dans des rôles de soutien. L'utilisation d'agents logiciels intelligents destinés à remplacer ces personnes pourrait permettre de réduire le nombre de seconds rôles et d'offrir à l'instructeur la possibilité de mieux encadrer la formation. Le présent rapport évalue un prototype de simulateur mis au point pour l'avion de patrouille maritime CP140 Aurora et doté d'agents logiciels intelligents qui jouent le rôle de navigateur tactique et d'opérateur de détecteur acoustique. On a réuni des équipages humains, des équipages d'agents intelligents et des équipages mixtes de personnes et d'agents intelligents pour remplir une mission anti-sous-marine simulée. On a consigné par écrit la qualité du travail accompli pendant la mission et noté les communications entre équipages afin de les évaluer, de manière à déterminer si les agents logiciels intelligents pouvaient accomplir des fonctions des membres d'équipage pris isolément et s'ils pouvaient assurer l'interaction nécessaire à l'entraînement à la coordination des équipages. Les constatations suivantes s'imposent : 1° les agents peuvent faire le travail des différents membres d'équipage;
2° l'interaction des agents avec les humains est suffisamment développée pour que ces derniers puissent s'acquitter de leurs tâches; 3° les relations des agents ne sont pas adaptées à l'entraînement à la coordination des équipages. On en conclut que le prototype convient à l'instruction individuelle, mais que la base de connaissance des agents doit tenir compte tout spécialement de la dynamique d'équipe avant de pouvoir servir à l'entraînement à la coordination des équipages.

P517317.PDF [Page: 7 of 35]

Executive summary

Use of intelligent software agents to play the role of supporting personnel in training simulations has the potential to reduce the requirement for support staff and increase the instructor's control of the simulation. The research presented here examines the suitability of intelligent software agents to aid training of individual crewmember skills and team skills. In this evaluation, human crews, intelligent agent crews, and mixed human-agent crews performed a simulated antisubmarine mission by a CP140 Aurora maritime patrol aircraft. Mission performance was recorded and crew communications were observed and rated to determine whether the intelligent software agents could perform individual crewmember functions and whether they could provide the interaction necessary for crew coordination training. The results indicate that (1) the intelligent software agents can perform individual crewmembers' functions adequately; and (2) the intelligent software agents did not interact in a way suitable for crew coordination training. The paper concludes with a discussion of the generalizability of the results and the growth potential of intelligent software agents in crew coordination training.

Sommaire

En faisant appel à des agents logiciels intelligents comme employés de soutien dans des simulations d'entraînement, on pourrait réduire l'effectif de ce côté et permettre à l'instructeur de mieux contrôler la simulation. La recherche dont il est question ici visait à déterminer si les agents logiciels intelligents étaient adaptés à l'acquisition de nouvelles compétences et à l'apprentissage de techniques d'équipe de la part des stagiaires. On a réuni, à des fins d'évaluation, des équipages humains, des équipages d'agents intelligents et des équipages mixtes de personnes et d'agents intelligents pour accomplir une mission anti-sous-marine simulée. On a consigné par écrit la qualité du travail accompli pendant la mission et on a noté les communications entre équipages afin de les évaluer, de façon à déterminer si les agents logiciels intelligents pouvaient accomplir des fonctions des membres d'équipage pris isolément et s'ils pouvaient assurer l'interaction nécessaire à l'entraînement à la coordination des équipages. Les constatations suivantes s'imposent : 1o les agents peuvent faire convenablement le travail des différents membres d'équipage; 2o les relations des agents ne sont pas adaptées à l'entraînement à la coordination des équipages. En conclusion, le mémoire comporte un exposé sur la généralisabilité des constatations de la recherche et sur les possibilités d'évolution des agents logiciels intelligents pour en faire des aides à l'entraînement à la coordination des équipages.

Grant, S.C. 2002. Assessing intelligent software agents for training maritime patrol aircraft crews. DCIEM TR 2001-036. Defence and Civil Institute of Environmental Medicine.

P517317.PDF [Page: 8 of 35]

Table of contents

Abstract i

Résumé i

Executive summary iii

Sommaire iii

Table of contents v

List of figures vii

List of tables vii

Acknowledgements viii

Introduction 1

Training requirement 1

Intelligent agents 2

Software technology 2

VCMS 3

The experiment 5

Method 5

Subjects 5

Apparatus 6

Procedure 9

Results 10

Mission performance 12

Communications analysis 14

Validity of VCMS 15

TACNAV's system evaluations 15

Discussion 16

Agents as OPFOR 17

P517317.PDF [Page: 9 of 35]

Agents for individual training 17
Agents for crew training 17
Comparision with other solutions 18

Recommendations for using intelligent agents for virtual role playing 19

References 20

List of symbols/abbreviations/acronyms/initialisms 22

P517317.PDF [Page: 10 of 35]

List of figures

Figure 1. Tactical Display During Scenario 1 11

Figure 2. Agent continues executing inappropriate plan 12

Figure 3. Distribution of grades assigned for each type of crew 13

Figure 4. Proportionate distribution of utterance types for each type of crew. 14

List of tables

Table 1. CCBC Categories Used to Analyze Utterances 8

Table 2. Responses To VCMS Evaluation Questionnaire 16

P517317.PDF [Page: 11 of 35]

Acknowledgements

I wish to acknowledge and thank the following people for their assistance in the prosecution of this research.

Ralf Kuehnel and David Chow for hardware and software support during extraordinarily trying circumstances.

407 Squadron and 14 Wing for the participation of Aurora aircrews in the experiment and for the use of their facilities during data collection.

Lt (N) Karl Boucher for his assistance as POC at CFB Greenwood during the data collection phase of this research.

Don Turner for data for assistance in data transformation.

P517317.PDF [Page: 12 of 35]

Introduction

In 1995, IntelAgent R&D Incorporated received a Defence Industrial Research (DIR) contract to investigate a new kind of simulation technology, *intelligent software agents* [1]. The investigation attempted to use intelligent agents for virtual role-playing, where computer programs emulate the behaviour of humans in the simulation. A follow-on DIR contract culminated in a prototype simulation of the CP140 Aurora that included agents that could play the role of the tactical navigator and acoustic sensor operator [2]. A successful development of this technology would be beneficial for the Aurora community. Project Management Office (PMO) Aurora observed that the use of agents could help mitigate the shortage of trained aircrews by allowing ground training to proceed without assembling an entire crew. Also, the normally quiet transit times during operational Aurora missions could be used as training time if the tactical crew in the Aurora could perform a simulated mission, interacting with agents playing the role of the flight deck crew while the real flight deck crew flew the aircraft. PMO Aurora therefore asked DCIEM to evaluate the prototype and make recommendations on the prospects of using intelligent agents to play the role of crewmembers in Aurora training simulations. This report documents how the evaluation was conducted, draws conclusions from the evaluation, and makes recommendations on the use of intelligent agents in training simulations.

Training requirement

The CP140 Aurora is the Canadian Forces' premier maritime patrol aircraft. Its long range, large payload, and diverse sensor suite allow it to perform a wide variety of missions. These missions included anti-submarine warfare (ASW), anti-surface warfare, search and rescue, and sovereignty patrols. Operation of this complex system necessitates a large crew that ranges from a normal complement of 10 to a maximum of 18.

Each crewmember has specific duties and must be skilled in operating a subset of the Aurora's mission systems. To aid in training the crew, a number of simulators are available. For example, the Cockpit Procedures Trainer, Acoustic Positional Trainer, and the Tactical Procedures Trainer simulate the crew stations of the flight deck crew, Acoustic Sensor Operators (ASOs), and Tactical Navigators (TACNAVs), respectively.

Beyond the requirement for individual proficiency, team skills are vital in determining mission success [3-5]. Crewmembers must work together as an integrated crew. Because they are interdependent, and must dynamically interact and adapt to the conditions that evolve during the flight, interactions amongst the TACNAV, ASOs, non-acoustic sensor operators (NASOs) and the navigator communicator (NAVCOMM) as well as the flight deck crew must be appropriate, co-ordinated and timely. The role of the TACNAV is particularly pivotal because the TACNAV leads the crew in tactical situations, integrating information and directing courses of action. To this end, the Aurora community has access to the Operational Mission Simulator (OMS). The OMS is a mock-up of the tactical crew compartment of the Aurora that allows the tactical crew to perform simulated anti-submarine warfare (ASW), anti-surface warfare (ASuW), and other missions.

To achieve the level of teamwork required for mission success demands that the crew have a common understanding, or overlapping mental models of their roles and the status of the mission [5-7]. Since "mental models are the mechanisms whereby humans generate descriptions of system purpose and form, explanations of system functioning and observed system states, and predictions of future system states" [8], it is vital that each crewmember's mental model be consistent with that of the other crewmembers. Accordingly, achieving and maintaining a mental model of the task that is consistent with one's crewmates' mental models is an important team skill.

Conducting team training in the OMS is subject to an obvious constraint; An Aurora crew is required. As a result, the time of already over-tasked aircrews may be further taxed. For instance, the need to train a new crewmember may occupy the time of at least six other crewmembers who need no further training. This cost will be incurred not only when new members join a squadron, but also in the upcoming Aurora upgrade program. In this program, individual Aurora crew stations will be upgraded, perhaps one aircraft at a time. As a consequence, crewmembers will have to be re-trained, one person at a time as their crew stations are modified. Furthermore, the requirement to assemble an entire crew means the training session is dependent on assembling all crew members. The absence of one of the crew can jeopardize the session unless a suitable replacement is found on short notice. Although the participation of a crew initially seems an inevitable part of team training, this constraint may evaporate in the light of emerging technology.

Intelligent agents

Software technology

Recent research in artificial intelligence is spawning a new technology called *intelligent agents*. Like expert systems, intelligent agents are programmed to solve problems within a specific domain. Intelligent agents are more than expert systems, however. They will act with autonomy, able to operate without intervention by a human. They are also both proactive, in that they spontaneously attempt to achieve their goals, and reactive, because they perceive their environment and change their behaviour accordlingly. Finally, they are social, interacting with humans, other agents, and conventional computer programs, to achieve their goals [9]. Thus they are well-suited to act in dynamic environments, where conditions are constantly evolving [10]

The nature of their design makes intelligent agents conducive to simulating the behaviour of humans. Already the simulation and training community is beginning to exploit this opportunity by using intelligent agents to play the role of individuals [11, 12] and entire teams [13, 14]. The motivation is to increase the controllability and affordability of simulations by reducing the number of supporting staff without losing their adaptability. The Virtual Crewmembers Simulator (VCMS) evaluated here attempts to go one step further by allowing a human student to obtain team skills by interacting with a team of intelligent agents.

VCMS

Denis Gagné [15, 16] recognized the potential for intelligent agents in meeting military training requirements. He proposed that the expertise, autonomy, interactivity, and adaptability of Aurora crew members might be sufficiently emulated by intelligent agents to allow them to act as virtual crew members in a simulation. This virtual role playing could then allow team training without the imposition of assembling an entire crew. On this insight Gagné and IntelAgent R&D developed the Virtual Crew Member Simulator (VCMS) to demonstrate that virtual role playing was capable of "providing both procedural and interpersonal process training to incomplete crews, while still maintaining vital team or crew concept interaction" [17].

At the foundation of the VCMS is a partial simulation of the tactical compartment of the CP140 Aurora. A subset of the sensors, systems, and weapons available to the TACNAV, ASO and NASO are available, and it is through these systems that the crew, human or virtual, interact with the virtual world. To focus development effort on the key topic of intelligent agents, the sensors, systems, and weapons are only provided to the degree required to perform the tracking and passive attack portions of a nuclear ASW mission. On top of the Aurora aircraft simulation are the intelligent agents. Intelligent agents are available to play the roles of the ASO and TACNAV and their rule bases allow them to localize and track submarines using passive acoustics. Rules covering the attack aspect of the mission were not available for evaluation.

Crew communication in the VCMS is accomplished through a simulation of the aircraft's intercommunication system (ICS). The ICS simulation is text based, where the human or agent crewmember assembles a message by selecting message fragments from a menu and typing the values for variables, such as speeds or sonobuoy numbers. The text output of the system is supplemented with voice production so that the human crewmember need not monitor the ICS text window. Other than that, the communication system is the same for both humans and agents.

The experiment

Although the ultimate goal of the intelligent agent technology is to redress a training deficiency, the VCMS is not a training device. Rather, it is an experimental simulator prototype. Consequently, the experiment assessed the ability of the agents to provide the appropriate crew interactions to support skilled team behaviour. The experiment investigated three related questions:

a. Does the limited functionality of the simulation provide a valid simulation of ASW? The validity of the simulation will be queried in two related ways. First, performance in the VCMS should show the effect of TACNAV experience, under the assumption that greater experience conveys greater skill. This is a refined version of the reverse transfer of training technique [18]. Second, a correlation should exist between performance in the OMS and performance in the VCMS. The top performing TACNAVs in the OMS should be the top performing TACNAVs in the VCMS.

b. Can a virtual crew member (VCM) demonstrate adequate individual skills? Aurora mission performance and TACNAV standards officer ratings will be used to compare the virtual TACNAV to human TACNAVs. Not only must the VCM be able to interact as part of the crew, it must be able to perform the tasks of a crew member.

c. Can Aurora crew members behave so that human crew members can exercise team skills? Crews establish and maintain their mental models through communication [4, 19, 20]. Therefore, the crux of the VCMS's potential as a team training device is the ability of the VCMs and human crew members (HCMs) to communicate in a way that allows mental models to be updated. This will be answered by examining the types of communications a human TACNAV has with a human and virtual ASO. The answer will indicate the sophistication level of Aurora team training possible using VCMs.

Method

The experiment used a within-subjects design, where the TACNAVs prosecuted an ASW mission first in the VCMS with a virtual crew and then another similar ASW mission in the OMS with a human crew. Additionally, an entirely virtual crew performed the same ASW mission in the VCMS.

Subjects

Eleven TACNAVs from the Canadian Forces 14 Wing acted as the experimental subjects. They were all qualified and currently serving as TACNAVs on operational crews. The squadron operations offices assigned the TACNAVs to the study, and the experimenter informed the TACNAVs verbally and in writing that the extent of their participation was up to them. They could terminate their experimental session at any point without repercussions.

Apparatus

Simulators

The VCMS is an unclassified system of several networked computers. The TACNAV's crew station consists of two computer monitors, a keyboard, and a mouse. One monitor mimics the computer monitor, or Multi-Purpose Display (MPD), that is the TACNAV's visual display in the Aurora. The other monitor is touch-sensitive, and displays an image of the keyset, or Integrated Control System (INCOS) used by the TACNAV in the Aurora. These interfaces were created using Virtual Prototypes VAPS product.

The TACNAV can enter most commands by pressing the buttons displayed on the touch-screen, just as he does in the aircraft. The numbers on the touch-screen INCOS do not work, however. The number pad on the keyboard is used instead. Finally, the mouse substitutes for the trackball that is used in the Aurora to indicate positions on the MPD.

The TACNAV's station in the VCMS cannot perform all the functions available in the Aurora, and some operate differently. This difference was inconvenient for the TACNAVs, but they had sufficient tools to perform the mission well and few complained about the deficiency during the debriefing. The range circle was the most missed function.

The Virtual Prototypes product STAGE was used as the scenario generator. The scenarios were developed initially by a TACNAV involved with the development of the VCMS, but modified for the purposes of the experiment. The course, speed, and position of the submarine were modified to increase the level of difficulty in the scenarios and to cause the tactical situation to depart from initial conditions more rapidly. From the outcome of this experiment, the scenarios presented some difficulty to the TACNAVs.

The human and virtual crewmembers in the VCMS communicate through a text window instead of an intercom as is done on the Aurora. By clicking the mouse pointer on the message fragment buttons and typing any required numbers on the keyboard, a human crew member constructs a message. The message is broadcast when the "send" button is clicked. The virtual crew members also construct messages from message fragments using software connections internal to the simulator and these too are broadcast. All broadcast messages are displayed in the text window and are also spoken aloud by a voice synthesis program.

The virtual crew members are programmed using SOAR. SOAR is software for creating artificially intelligent systems and it is an expression of a unified theory of human cognition [21, 22]. One ASO agent and three TACNAV agents (standard, novice, and slow) are available. They are programmed using 369 and 533 rules for the ASO and standard TACNAV, respectively. The agents interact solely with the sensor and systems simulations. They do not

have privileged access to scenario data and have no capabilities or knowledge that are unavailable to human users of the VCMS. The slow TACNAV agent had the same number of rules, but executed them more slowly. The novice TAVNAV agent lacked rules that allowed it to use single sonobuoys to fill spaces between existing sonobuoy patterns. The novice TACNAV agent would fill the gaps between patterns using new patterns. Their rule bases allowed them to localize and track submarines using passive acoustics. Rules covering the attack aspect of the mission were not available for evaluation.

The OMS is a weapon systems trainer that is amock-up of the tactical compartment of the Aurora. Crew stations for the TACNAV, NAVCOM, ASO1, ASO2, NASO1, and NASO2 are present. The instructor console has a workstation for the pilot and an instructor, who controls the scenario and acts as the NASO3 when required. The scenarios are fully controllable, and the simulator has record and playback capabilities. The simulator is capable of simulating much more of the mission profile than the VCMS, so creating a scenario comparable to the ones used in the VCMS was straightforward.

Crew Communications Behaviour Categories

Two judges used the Crew Communications Behaviour Categories (CCBC) to characterize the communications amongst crewmembers in the study. The CCBC is a rating system derived from theory of the limitations on human information processing and mental models and implemented specifically to address teamwork in CF aircrews [23] [24]. The subset of the CCBC categories used in this evaluation is presented in Table 1, while the full set of categories is contained in Appendix A. The keying of the microphone in the OMS or the contents of the text string in the VCMS defined the beginning and end of an utterance. An utterance could be assigned to one or more categories.

P517317.PDF [Page: 18 of 35]

Table 1. *CCBC Categories Used to Analyze Utterances*

CATEGORY NAME	DEFINITION	EXAMPLE
Awareness	Declarative statements concerning present states	"On heading for E1"
Preplanning	Explicit statements of future intentions or plans	"I'm putting in an intercept I'd like you to stay back so you can come down for the attack "
Task Prioritization	Overt direction or request for direction concerning task prioritization	"Pilot, we want E1 then E2"
Criticism	Negative statements concerning performance	"All you have to do is ask, eh?" (sarcastic)
Positive Reinforcement	Positive statements concerning performance	"Well put"
Crew Coordination	Explicit statements concerning the apportioning of tasks between crew	"Roger that's fine go to the next one I'll destroy this one "
Cross-Checking	Explicit noting of deviations with respect to another crewmember's actions	"How are you doing on that one?"
Implications	Explicit recognition of the consequences of present system states	" Buoy 19 is going to die and we'll lose attack criteria"
Queries	Questions or soliciting information	"Is he showing higher on 9 and 10?"
Proactive	Explicit proffering of information, not in response to a query	" Tac, do you want a time slice right now?"
Directive	Commands or directives	"OK pilot, let's descend to 300 feet"
Input	Information provided in response to a query	" Your strongest contact is 10 "
Acknowledgement	Statements to acknowledge actions or statements	"Roger"

The TACNAVs completed a simulator system evaluation questionnaire. This instrument uses both Likert scales and open-ended questions to elicit their opinions of the simulator. A copy of the evaluation questionnaire is contained in Appendix B.

The Standards and Training Officer responsible for evaluating TACNAVs served as the subject matter expert. He rated mission performance using the TACNAV B Cat Checkride Assessment Form, an instrument used to qualify TACNAVs. The form provides a rating scale, from 1 to 5, for each task performed by the TACNAV. Higher numbers denote superior performance

and the form provides behavioural anchors for assigning ratings. The distribution of ratings generates a letter grade as well as a pass or fail assessment. A TACNAV is deemed to fail when fewer than 80% of the tasks are rated 4 or 5, when any task is rated 1 or 2, or where the deployment of weapons or search stores is rated 3 or less. The subject matter expert completed only the sections of the form relevant to the scenarios used. These were the ratings for following enabling objectives: Provide Tactical Steering, Deploy Search Stores, Deploy Weapons, Maintain Sonobuoy Reference System Plot Stabilization, Understand Doppler, Conduct ASW Nuclear. The subject matter expert also utilized the section provided for open-ended comments. A copy of the form is presented in Appendix C.

Procedure

Ten TACNAVs first completed two ASW missions in the VCMS working with an agent ASO. One TACNAV was unable to complete the VCMS missions due to a failure in the VCMS. Next, they completed another ASW mission in the OMS working with their usual human crewmates. The time between the VCMS session and the OMS session was approximately 12 weeks.

At the beginning of the VCMS session the experimenter explained to the TACNAVs that their participation was sought to assess the utility of the agent technology, and not to assess their skills. After the TACNAVs consented to participate in the study they were familiarized with the VCMS system. The limitations of the VCMS were explained and they practiced using the touchscreen controls and interacted with the ASO via the ICS until they were comfortable with their knowledge of the system, normally about 5 minutes. Afterward, the experimenter briefed the TACNAVs on the first scenario and answered any questions about the VCMS and the scenario.

In the scenario briefing, the TACNAVs were told the location of the scenario and that their mission was to localize and track a hostile submarine known to be somewhere in their area of responsibility. They were told that there were no other aircraft, ships, or submarines in the area, only themselves and the target. The rules of engagement at the start of the scenario did not allow them to attack the submarine immediately, but they were told that authorization might come during the scenario.

The TACNAVs then began the first scenario. It begins with a pattern of 12 sonobuoys in the water and the submarine located so that one sonobuoy quickly obtained direct path contact with the submarine. The submarine followed a preset course away from the initial pattern of sonobuoys, changing its speed and heading periodically to make tracking challenging and make difficult the planning and placement of new patterns of sonobuoys. The initial contact was sufficient to enable the ASO to provide a rough bearing to the submarine, but a new pattern of sonobuoys had to be quickly laid before the submarine moved out of range. Twenty five minutes into the scenario the experimenter informed the TACNAVs that they were authorized to attack the submarine. The scenario ended when the TACNAV hit the submarine with a torpedo

or the when 60 minutes elapsed. All TACNAVs were able to hit the submarine before the deadline.

The second scenario was then briefed and played out. The second scenario was similar to the first, except that it began with a pattern of two sonobuoys and the authorization to attack came in 20 minutes. Obviously, the position and course of the submarine were different. All TACNAVs successfully attacked the submarine before the end of the scenario.

At the end of the VCMS session, the TACNAVs completed a questionnaire soliciting their opinions on various aspects of the VCMS.

In the OMS session, the TACNAVs performed one ASW mission that was similar to the ones they performed in the VCMS. The whole crew was briefed on the scenario and the scenario continued until either the submarine was sunk or four hours elapsed. Data collection ended, however, when the first torpedo hit the submarine or the four-hour time limit was reached. All crews were able to successfully attack the submarine before the deadline.

The all-agent crew also performed the two VCMS scenarios. Because the agents did not have the rule base required to carry out an attack, they continued to track the submarine until the time limit expired.

Results

Analysis of the data from the VCMS began with replaying the recorded missions by all the crews and observing them for notable characteristics. This revealed both similarities and differences between the human and agent TACNAVs. Figure 1 is a view of the tactical situation as it appeared on an agent TACNAV's CRT during the first scenario. Moments before this point in the scenario the ASO obtains contact with the submarine just long enough for the TACNAV to estimate its course. When contact is then lost and the submarine changes course from 015 to 090, the behavior of the agent TACNAV is much like that of the human TACNAVs. They both rely on the existing pattern of sonobuoys to cover possible movement to the south, and they plan another pattern of sonobuoys ahead of the submarine based on the assumption that the submarine maintains the same course.

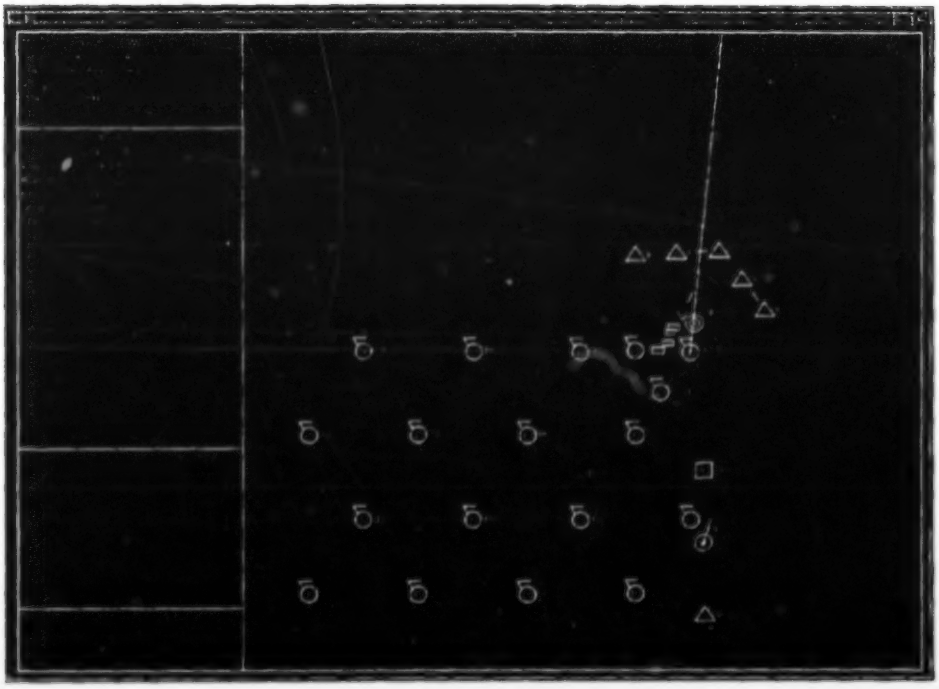

Figure 1. *Tactical Display During Scenario 1 Tactical display during scenario 1, as performed by an agent TACNAV Sonobuoys in the water are represented by δ = sonobuoy in the water, Δ = planned sonobuoy positions; v = estimated position of the submarine, ❶ = true position of the submarine*

The granularity of their actions differ, however. The agent plans a five buoy wedge pattern (Figure 2), based on one of the preset patterns available in the aircraft mission system, whereas human TACNAVs typically drop less than a complete pattern and thereby conserve buoys and freedom of movement. Furthermore, although contact is regained shortly after this point in the scenario, and the new course of the submarine obtained, the agent remains committed to completing the previously planned operation, the five buoy wedge pattern to the north, despite its updated knowledge of the submarine. Human TACNAVs, of course, are quick to interrupt their previous pattern at whatever stage of completion and continuously adapt their behavior.

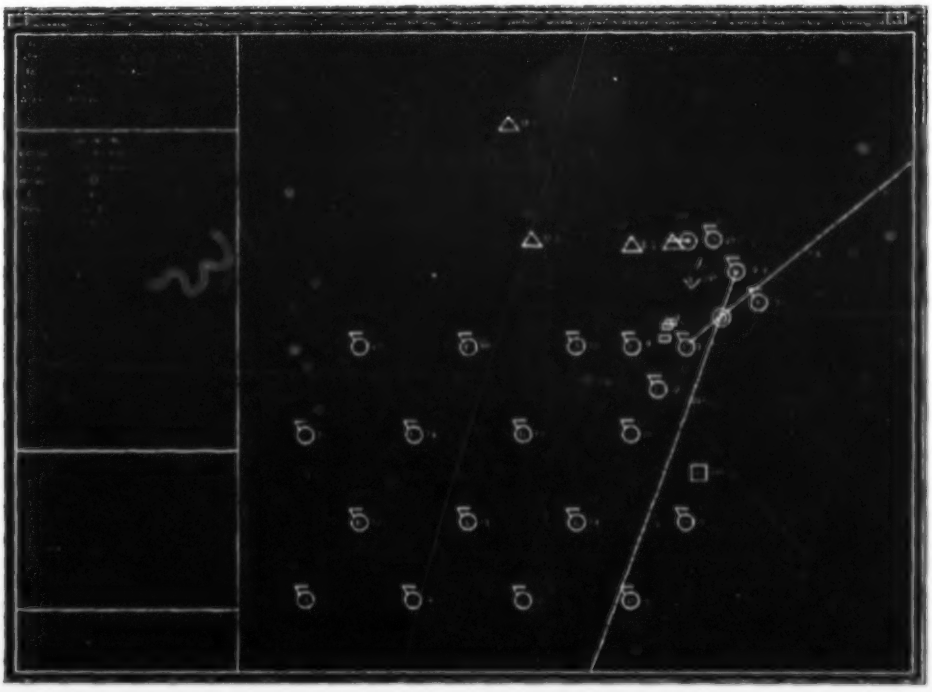

Figure 2. *Agent continues executing inappropriate plan*

Mission performance

Figure 3 presents the ratings assigned by the Standards Officer to each of the tasks performed by the TACNAVs in the VCMS. Considerable overlap in the distribution of ratings earned by the human-agent and agent-agent crews is apparent, although the higher ratings tend to be earned by the human-agent crews. Using a Mann-Whitney U test comparing the modal ratings earned by the human-agent crews and the agent-agent crews is inconclusive because of the large number of ties in the data, $U = 7.00$, $p > .1$. This is unavoidable due to the deterministic nature of the agent crews, the limited number of different agent crews, and the limited behavioural repertoire of the agent crews.

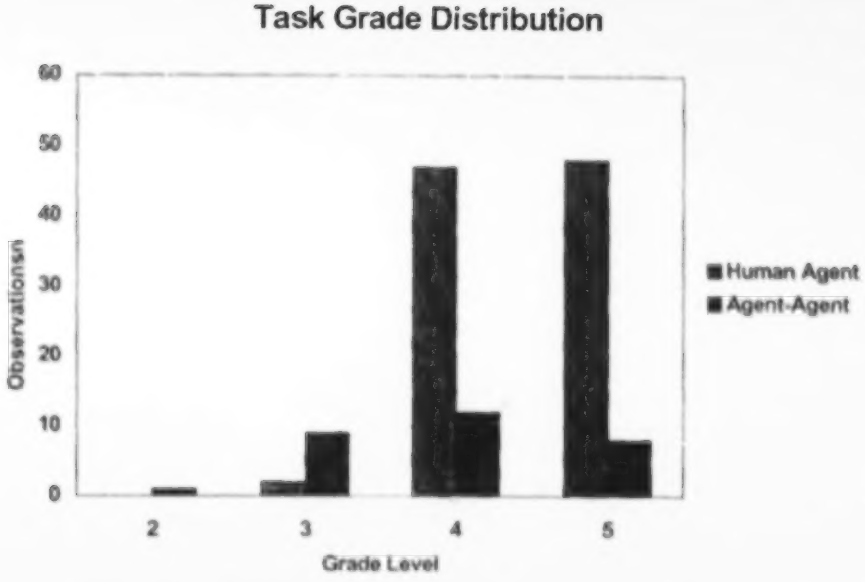

Figure 3. *Distribution of grades assigned for each type of crew*

Turning to the pass/fail data, the standard TACNAV and four human TACNAVs received failing grades in the first scenario. The standard agent failed on three criteria: scoring a 2 on Provide Tactical Steering, scoring a 3 on Deploy Search Stores, and scoring less than 4 on more than 80% of the tasks. Many of the Standards Officer's comments focussed on the standard agent's failure to prepare for possible future developments in the scenario. The standard agent positioned the aircraft too far away from the submarine to be able to react promptly and waited until contact was nearly lost before placing new sonobuoys. The Standards Officer also criticized the standard agent for wasting sonobuoys by completing patterns that were found to be poorly positioned during the course of their construction. The slow TACNAV agent failed for scoring a 3 on Deploy Search Stores and for scoring less than 4 on more than 80% of the tasks. Again, the Standards Officer criticized the slow agent for waiting too long until contact was nearly lost before dropping new sonobuoys. The Standards Officer noted that the slow agent TACNAV was behaving unreasonably because when it lost contact with the submarine, it continued to place patterns beyond existing patterns that the submarine never penetrated. The novice TACNAV failed for scoring less than 4 on more than 80% of the tasks. The Standards Officer indicated that the novice agent TACNAV was too slow in placing new sonobuoys, positioned the aircraft poorly, and placed sonobuoy patterns on the basis of little information. Human TACNAV number 1 failed for scoring 3 on Deploy Search Stores. The

Standards Officer criticized the human TACNAV for being slow to position sonobuoy patterns. This comment is attributable to a slowdown in the TACNAV's user-interface observed when TACNAV number 1 performed the scenario. Only one TACNAV failed the second scenario. Human TACNAV number 12 failed for scoring a 3 on Deploy Weapons. The Standards Officer's comments indicated that human TACNAV number 12 failed for launching a weapon before achieving launch parameters.

Communications analysis

The judges independently categorized the crew's utterances using the CCBC. Agreement between the judges was measured using the κ statistic. The result, $\kappa = 0.65$, indicated that there is "substantial" agreement between the judges [25].

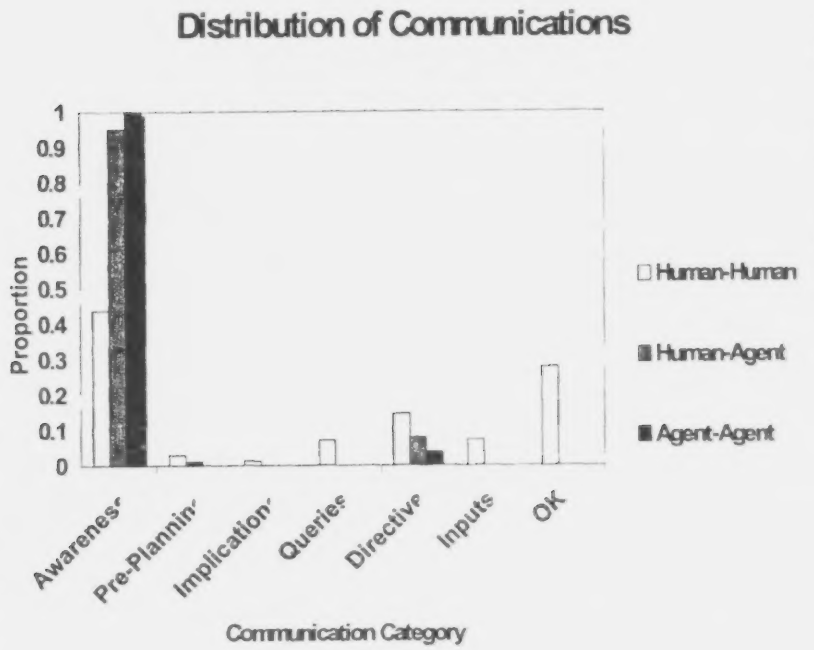

Figure 4. *Proportionate distribution of utterance types for each type of crew.*

The distributions of utterance types with a frequency of at least 1% are presented in Figure 4 for each of the three different crew types (human-human, human-agent, and agent-agent). Only the human-human crews used the utterance types not appearing in the figure (task prioritization, positive reinforcement, cross-checking, crew coordination, criticism, and proactive). It is apparent that the communications

involving agents were largely confined to awareness statements and directives. Not surprisingly, when χ^2 goodness of fit tests were used to compare the utterance distributions of each of the human-agent and agent-agent crews to the utterance distribution that characterized human-human crews, the result of each test was significant, χ^2 (df=12) > 26.22, p < 01.

Validity of VCMS

The correlation between the number of months a TACNAV held TACNAV qualification and four different measures of performance in the VCMS did not reveal a consistent pattern. The correlation of experience with the mean number of sonobuoys expended was reliable, r = .86, p < .10, df = 6. The correlations of experience with mean time to hit the submarine, torpedo drop error, and the mean grade in the VCMS, however, were not statistically reliable. Correlating experience with performance in the OMS produced non-significant results on all variables: sonobuoys expended (r = .33), time to hit (r = .05), and torpedo drop error (r = .48).

A different validity check was performed by correlating performance in the OMS with performance in the VCMS. The correlations of the number of sonobuoys expended (r = .49), torpedo drop error (r = -.34), and time to hit the submarine (r = .09) in one simulator did not reliably predict the corresponding value in the other simulator.

TACNAV's system evaluations

The responses of the TACNAVs are presented in Table 2. Their opinion of the VCMS in general was non-committal, endorsing mid scale values on questions 1 and 11. When question 2 asked for their "gut feeling" of whether the simulator felt right, they tended to respond negatively. This dissatisfaction was not due to difficulties in sending messages to the agent ASO. On question 3 the TACNAVs reported the text ICS as easy to use. Question 4 revealed that receiving messages on the ICS was less satisfactory. It is not a surprise, therefore that the TACNAVs reported in question 5 that the ICS affected the way they interacted with the agent ASO and question 9 refined this to indicate that the interaction was different from what the TACNAVs experienced with their own crews. Question 6 indicated the ICS had other effects on the way some, but not all, TACNAVs performed the mission. When asked in question 7 about the competence of the agent ASO, there was little agreement amongst the TACNAVs . Nevertheless, in question 8 the TACNAVs showed that the difference between the agent ASO and a human ASO caused them to perform the mission differently. The mockup of the TACNAV's INCOS and MPD was certainly acceptable, with the responses to question 10 being favorable.

P517317.PDF [Page: 26 of 35]

Table 2. Responses To VCMS Evaluation Questionnaire

SUBJECT	QUESTION										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
01	2	7	1	7	7	7	6	7	7	4	6
02	Simulator failed so no questionnaire data collected										
03	5	6	2	3	6	3	4	7	7	3.5	3
04	3	3	3	5	5	2	3	2	2		
05	Agent TACNAV										
06	Agent TACNAV										
07	3	4	3	2	2		4	3	6	3	2
08	4	4	2	6	7	5	6	3	6	6	4
09	Agent TACNAV										
10	3	5	2	4	6	6	4	6	6	2	3
11	3	4	3	6	6	6	3	5	5	2	3
12	3	6	3	5	7	5	5	7	7	5	4.5
13	3	3	2	3	2	1	1	4	5	2	3
14	5	6	2	5	7	5	7	7	7	2	3
Mean	3.40	4.80	2.30	4.60	5.50	4.44	4.30	5.10	5.80	3.28	3.50
s.d.	0.92	1.33	0.64	1.50	1.86	1.89	1.68	1.87	1.47	1.40	1.11

Discussion

Observing the performance of human-agent crews and agent-agent crews in contrast to human-human crews revealed both strengths and weaknesses of the agent paradigm that suggest how the technology might best be used.

Figure 1 is a view of the tactical situation as it appeared on an agent TACNAV's MPD during the first scenario. Moments before this point in the scenario the ASO obtains contact with the submarine just long enough for the TACNAV to estimate its course. When contact is then lost and the submarine changes course from 015 to 090, the behavior of the agent TACNAV is much like that of the human TACNAVs. They both rely on the existing pattern of sonobuoys to cover possible movement to the south, and they plan another pattern of sonobuoys ahead of the submarine based extrapolation from the available data.

The agents and human differ in the decomposability of their actions, however. The agent plans a wedge pattern of five sonobuoys, using the one of the preset patterns available in the aircraft mission system to do so. The humans, on the other hand, typically drop fewer than a full pattern of sonobuoys, thereby conserving search stores and freedom of movement. Furthermore, when contact is regained shortly after this point in the scenario, the new westerly course of the submarine obtained, yet the agent remains committed to completing the

previously planned pattern to the north. Human TACNAVs, of course, are quick to interrupt their previous pattern at whatever stage of completion and immediately adapt their behavior to the new data.

Agents as OPFOR

The performance of the agent-agent crews was reasonably good. They earned passing grades in one scenario and failing grades on the second. Furthermore, two of the agent-agent crews actually outscored a human-agent crew in one of the scenarios. Such a result from an early prototype is encouraging and suggests that competent performance from simulated crews can be obtained via the intelligent agent paradigm. Although the agents are inferior to human crews, they could be profitably used to playing the role of an opposing force. A submarine crew training in a simulator might benefit from operating against an agent-controlled MPA, for example.

Agents for individual training

The passing grades obtained by the human-agent crews demonstrate that the agent ASOs were competent enough to allow human TACNAVs to exercise their individual skills. The rudimentary ICS available at this stage in the VCMS development provided enough communications between the human and agent for them to jointly achieve the mission objectives. This reveals that intelligent agent technology can support the task performance and interactivity required for training individual skills.

Agents for crew training

The communications analysis revealed clear differences between the human-human crews and the crews with one or two agents. Some of the discrepancy arises from rather minor shortcomings, such as the absence of acknowledgements in the crews containing agents. Although acknowledgements are important communications acts, their absence merely reflects the prototype status of the VCMS and they could easily be added to the agent's behavioral repertoire.

Other discrepancies pose a greater challenge. Only the human-human crews made crew coordination and cross-checking utterances. The immediate, facile explanation is that the ICS interface did not offer these kinds of utterances, but that is not the root cause. Human TACNAVs rarely made preplanning utterances to agents, even though there was such an utterance (the attack briefing) available for broadcast on the ICS. Agent TACNAVs never made preplanning utterances. The ultimate reason for the dearth of certain types of utterance is that the agents are incapable of *metacognition*. Their knowledge bases do not contain information regarding the mental operations and limitations of the crewmembers so they did not make metacognitive statements. The agents also do not have knowledge bases regarding their own "cognition", so they do not benefit themselves from metacognitive utterances. For example, preplanning in human crews cues memory, directs attention, and manages workload, but it has no effect on the agents. Giving an attack briefing to an agent was a therefore a task that consumed the TACNAV's time and cognitive resources, but did not payoff

in better crew performance. It is not surprising, therefore, that the human TACNAVs would skip the attack briefing and work on other tasks that could affect the course of the scenario. When they did give the attack briefing, it appeared to be *pro forma*. If the agents in their present form were to be used for crew coordination training the potential for negative training is obvious.

The inability for these agents to support some of the kinds of communications that are central to effective teams limits their usefulness for training team skills. It is important to note however, that this shortcoming does not apply in principle to all agents, or even these agents in particular. Dedicated programming effort may remedy this shortcoming. The point is simply that the solution does not automatically or elegantly follow from the intelligent agent paradigm in the same way that other attributes do. To meaningfully use utterances dealing with metacognition will require the agent to have a model of the current mental status of the other crewmembers. This constitutes an additional knowledge base, one dealing with the cognitive demands placed on the crewmembers. Coupled with this is a need for linguistic competence to deal with the complex statements regarding metacognition. These statements are more difficult for artificially intelligent systems of any kind to handle than directives and awareness statements owing in part to their more varied syntax and less constrained vocabulary.

Comparision with other solutions

In addtion to the OMS, the Aurora community also has access to the Tactical Procedures Trainer (TPT). The TPT is a stand-alone simulator that provides a TACNAV with an INCOS, MPD, and the services of an ASO. The TPT's ASO is a more mature piece of software than the agent ASO in the VCMS and can perform more of the ASOs duties at a higher level of competency. The result is a simulator that allows TACNAVs to exercise a broad range of their individual skills in user-definable scenarios. Its functionality exceeds the current VCMS and the simulation runs on a very modest computer system. Although no formal evaluation has been performed, TACNAVs and instructional staff of 14 Wing feel that it provides an effective learning platform for new TACNAVs.

The TPT's design goals were not as ambitious as those of the VCMS. It was not intended to provide the interactivity that the VCMS attempts to provide for the training of team skills. As a result, the ASO is fully integrated into the system and the TPT is a stand-alone system that will not communicate with other computers. This means that the TPT cannot be networked with other simulators to share data. Also, a human ASO cannot replace the TPT's ASO and interact with a human TACNAV. Furthermore, the integration of the ASO means that changing the behaviour of the TPT's ASO independently of the rest of the simulation may be more difficult than doing so to an ASO agent.

Recommendations for using intelligent agents for virtual role playing

- The TPT currently delivers valuable individual skills training to TACNAVs. Replacing or maintaining this capability is not sufficient reason for further pursuit of agent technology by PMO Aurora.

- Current intelligent agents can perform tasks well enough to play the role of individual actors in simulations, as demonstrated here and in the work of others [11-14]. The Defence Research and Development Canada (DRDC) should continue working on agent technology for virtual role-playing. Research should particularly address the modelling of human characteristics where relevant, as opposed to subject-matter expertise.

- Agents are best used where their interactions with humans will occur through either standardized communications protocols or through non-linguistic means (e.g. agents as OPFOR). Humans and agents should not be mixed where their interaction is unconstrained and not formalized; rather, greatest success will be enjoyed when agents represent all of the actors within a boundary defined by formal communications procedures. PMO Aurora should not pursue agent technology for training team skills in the near to mid term. DRDB should work on increasing our capability of providing human – agent interaction and provide guidance for implementation.

- DRDB should be particularly receptive to industrial interest in human behaviour modelling for real-time simulation. The CF, DND, and Canadian industry is reaping the benefits of past support of the Canadian simulation industry by the DRDB. Similar support for companies commercializing human behaviour modelling will provide an opportunity to transfer DRDB knowledge of human behaviour and thereby increase the value of simulations, and improve the products of Canadian industry.

P517317.PDF [Page: 30 of 35]

References

1. IntelAgent R&D Inc, Virtual Interactive Role Playing For Military - DIR Proposal No. 199 - (Part of the Defence Industrial Research (DIR) Program of DND), in Lochlan Magee, P.W.a.G.S. Canada, Editor. 1995.

2. IntelAgent R&D Inc, NOVA2: Virtual Interactive Role Playing for Military Training - DIR Proposal No.: 199 Extension - (Part of the Defence Industrial Research (DIR) Cost-Shared Program of the Department of National Defence, in Stuart Grant, P.W.a.G.S. Canada, Editor. 1996.

3. Hill, G.W., Group Versus Individual Performance: Are N + 1 Heads Better Than 1? Psychological Bulletin, 1982. 91(3): p. 517-539.

4. Kanki, B.G. and M.T. Palmer, Communication and Crew Resource Management, in Cockpit Resource Management, E.L. Wiener, B.G. Kanki, and R.L. Helmreich, Editors. 1993, Academic Press: San Diego, CA. p. 99-137.

5. Salas, E., et al., Toward an Understanding of Team Performance and Training, in Teams: Their Training and Performance, R.W. Swezey and E. Salas, Editors. 1992, Ablex Publishing Corporation: Norwood, NJ. p. 3-29.

6. Cannon-Bowers, J.A., E. Salas, and S. Converse, Shared Mental Models in Expert Team Decision Making, in Individual and Group Decision Making, N.J. Castellan, Editor. 1993, Lawrence Erlbaum Associates: Hillsdale, NJ. p. 221-246.

7. Orasanu, J.M., Shared Mental Models and Crew Decision Making. 1990, Princeton University: Princeton, NJ. p. 21.

8. Rouse, W.B. and N.M. Morris, On looking into the black box: Prospects and limits in the search for mental models. Psychological Bulletin, 1986. 100(3): p. 349-363.

9. Wooldridge, M. and N.R. Jennings, Intelligent Agents: Theory and Practice. The Knowledge Engineering Review, 1995. 10(2): p. 115-152.

10. Bradshaw, J.M., An Introduction to Software Agents, in Software Agents, J.M. Bradshaw, Editor. 1997, MIT Press: Cambridge MA. p. 3-46.

11. Glenn, F., et al. A Simulation Model of the Operator of an Air Defense Weapons System. in Seventh Conference on Computer Generated Forces and Behavioral Representation. 1998. Orlando, FL.

12. Howard, M. and C. Lee. Architecture of a Generic Command Entity. in Seventh Conference on Computer Generated Forces and Behavioral Representation. 1998. Orlando, FL.

13. Hill, R.W., et al. Soar-RWA: Planning, Teamwork, and Intelligent Behavior for Synthetic Rotary Wing Aircraft. in Seventh Conference on Computer Generated Forces and Behavioral Representation. 1998.

14. Zachary, W., et al. Simulating Behavior of Tactical Operators and Teams Using COGNET/GINA. in Seventh Conference on Computer Generated Forces and Behavioral Representation. 1998.

15. Gagné, D., et al. Aurora: A Multi-Agent Prototype Modelling Crew Interpersonal Communication Network. in Department of National Defence Workshop on Advanced Technologies in Knowledge Based Systems and Robotics. 1993. Ottawa, Ontario.

16. Gagné, D., et al. Aurora: A Multi-Agent Modelling Crew Interpersonal Communication Network. in The 5th Computer Generated Forces and Behavioral Representation Conference. 1995. Orlando, Florida.

17. Gagné, D., DIR Proposal Submission NOVA: Virtual Interactive Role Playing for Military Training. 1995, IntelAgent R&D Inc: Victoriaville, Quebec.

18. Rolfe, J.M. and P.W. Caro, Determining the training effectiveness of flight simulators: Some basic issues and practical developments. Applied Ergonomics, 1982. 4(December): p. 243-250.

19. Kanki, B.G., S. Lozito, and H.C. Foushee, Communication Indices of Crew Coordination. Aviation, Space, and Environmental Medicine, 1989. 60: p. 56-60.

20. Foushee, H.C., The Role of Communications, SocioPsychological, and Personality Factors in the Maintenance of Crew Coordination. Aviation, Space, and Environmental Medicine, 1982. 53(11): p. 1062-1066.

21. Newell, A., Unified Theories of Cognition. 1990, Cambridge, MA: Harvard.

22. Rosenbloom, P.S., J.E. Laird, and A. Newell, The Soar Papers: Readings on Integrated Intelligence. Vol. 1 and 2. 1993, Cambridge, MA: MIT Press.

23. Thompson, M.M., Human Factors of CC-130 Operations. Volume 2: Crew Performance Measures. 1998, Defence and Civil Institute of Environmental Medicine: Toronto, ON.

24. Banks, R.D., et al., Human Factors Study of CC-130 Operations. 1996, Defence and Civil Institute of Environmental Medicine: Toronto, ON.

25. Landis, J.R. and G.G. Koch, The Measurement of Observer Agreement for Categorical Data. Biometrics, 1977. 33(1): p. 159-174.

List of symbols/abbreviations/acronyms/initialisms

ASO	Acoustic Sensor Operator
ASuW	Anti-surface warfare
ASW	Anti-submarine warfare
CCBC	Crew Communication Behaviour Categories
CF	Canadian Forces
DCIEM	Defence and Civil Institute of Environmental Medicine
DIR	Defence Industrial Research
DND	Department of National Defence
DRDC	Defence Research and Development Canada
ICS	Intercommunication system
INCOS	Integrated Control System
MPD	Multi-purpose display
NASO	Non-acoustic sensor operator
OMS	Operational Mission Simulator
OPFOR	Opposing Force
PMO	Program Management Office
TACNAV	Tactical Navigator
VCMS	Virtual Crew Member Simulator

DOCUMENT CONTROL DATA SHEET

1a. PERFORMING AGENCY	2 SECURITY CLASSIFICATION
DCIEM	UNCLASSIFIED Unlimited distribution -
1b PUBLISHING AGENCY DCIEM	

3 TITLE

(U) Assessing Intelligent Software Agents for Training Maritime Patrol Aircraft Crews

4 AUTHORS

Grant, S.C.

5 DATE OF PUBLICATION	6 NO OF PAGES
December 1 , 2001	34

7 DESCRIPTIVE NOTES

8 SPONSORING/MONITORING/CONTRACTING/TASKING AGENCY

Sponsoring Agency
Monitoring Agency
Contracting Agency
Tasking Agency PMO Aurora

9. ORIGINATORS DOCUMENT NO	10 CONTRACT GRANT AND/OR PROJECT NO	11 OTHER DOCUMENT NOS.
Technical Report TR 2001-036	6b	

12 DOCUMENT RELEASABILITY

Unlimited distribution

13 DOCUMENT ANNOUNCEMENT

Unlimited announcement

14 ABSTRACT

(U) Training simulators often require the participation of several people to play the role of supporting players in the simulated operation. Use of intelligent software agents to play the role of these personnel has the potential to reduce support staff and increase an instructor's control of training. This report evaluates a simulator prototype developed for the CP140 Aurora maritime patrol aircraft that incorporated intelligent software agents to play the roles of the Tactical Navigator and an Acoustic Sensor Operator. Human crews, intelligent agent crews, and mixed human-agent crews performed a simulated antisubmarine mission. Mission performance and crew communications were recorded and rated to determine whether the intelligent software agents could perform individual crewmember functions and whether they could provide the interaction necessary for crew coordination training. The results indicate that (1) agents can perform individual crewmembers' functions; (2) agent interaction with humans is sufficient to allow humans to perform their own tasks; and (3) the agents did not interact in a way suitable for crew coordination training. It is concluded that the prototype is suitable for supporting individual training, but the agents' knowledge base must explicitly address team dynamics if crew coordination training is to be supported.

15 KEYWORDS, DESCRIPTORS or IDENTIFIERS

(U) intelligent agents; Aurora; constructive simulation

Defence R&D Canada
is the national authority for providing
Science and Technology (S&T) leadership
in the advancement and maintenance
of Canada's defence capabilities.

R et D pour la défense Canada
est responsable, au niveau national, pour
les sciences et la technologie (S et T)
au service de l'avancement et du maintien des
capacités de défense du Canada.

#517317
CA020611

DEFENCE R&D DÉFENSE

www.drdc-rddc.dnd.ca